



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>6</sup> :

G06T 5/00

A1

(11) Numéro de publication internationale:

WO 99/40539

(43) Date de publication internationale:

12 août 1999 (12.08.99)

(21) Numéro de la demande internationale: PCT/FR99/00176

(22) Date de dépôt international: 28 janvier 1999 (28.01.99)

(30) Données relatives à la priorité:

98/01308

4 février 1998 (04.02.98)

FR

(71) Déposant (pour tous les Etats désignés sauf US): UNIVERSITE  
PIERRE ET MARIE CURIE (PARIS VI) [FR/FR]; 4, place  
Jussieu, F-75252 Paris Cedex 05 (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (US seulement): FAUDEMAY, Pascal  
[FR/FR]; 33, rue Saint-Amand, F-75015 Paris (FR). DU-  
RAND, Gwenaél [FR/FR]; 4, rue Dussoubs, F-75002 Paris  
(FR).(74) Mandataires: GUTMANN, Ernest etc.; Ernest Gutmann-Yves  
Plasseraud S.A., 3, rue Chauvau-Lagarde, F-75008 Paris  
(FR).(81) Etats désignés: BR, CA, IL, IN, JP, US, brevet européen (AT,  
BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,  
MC, NL, PT, SE).

Publiée

*Avec rapport de recherche internationale.**Avant l'expiration du délai prévu pour la modification des  
revendications, sera republiée si des modifications sont  
reçues.*

(54) Title: METHOD FOR SPATIAL SEGMENTATION OF AN IMAGE INTO VISUAL OBJECTS AND APPLICATION

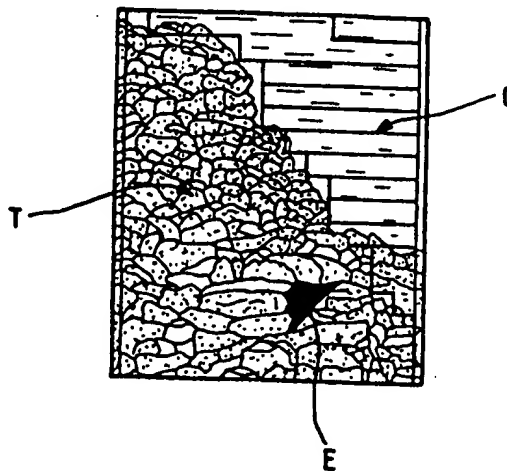
(54) Titre: PROCEDE DE SEGMENTATION SPATIALE D'UNE IMAGE EN OBJETS VISUELS ET APPLICATION

(57) Abstract

The invention concerns a method for spatial segmentation of images into visual objects to obtain objects having semantic significance comprising steps which consists in dividing the image into initial regions, fusing the neighbouring initial regions according to certain similarity functions and obtaining enlarged images, then repeating said phase with the remaining initial regions and the enlarged regions until there is no possibility of further fusion. The invention also concerns the use of said method for characterising, classifying, storing, detecting objects, scenes, planes, groups of planes, movements, in multimedia documents, particularly of video.

(57) Abrégé

L'invention porte sur un procédé de segmentation spatiale d'images en objets visuels pour obtenir des objets ayant une signification sémantique comportant des étapes de partition d'images en régions initiales, de fusion de régions initiales voisines selon certaines fonctions de similarité et l'obtention de régions élargies, puis une itération de cette phase avec les régions initiales restantes et les régions élargies jusqu'à ce qu'aucune fusion ne soit plus possible. L'invention porte également sur l'utilisation du procédé à la caractérisation, à la classification, au stockage, à la détection d'objets, de scènes, de plans, de groupes de plans, de mouvements, dans des documents multimédia en particulier de vidéo.



# UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce			TR	Turquie
BG	Bulgarie	HU	Hongrie	ML	Mali	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MN	Mongolie	UA	Ukraine
BR	B Brésil	IL	Israël	MR	Mauritanie	UG	Ouganda
BY	Bélarus	IS	Islande	MW	Malawi	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	MX	Mexique	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Pays-Bas	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NO	Norvège	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	NZ	Nouvelle-Zélande		
CM	Cameroun			PL	Pologne		
CN	Chine	KR	République de Corée	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Roumanie		
CZ	République tchèque	LC	Sainte-Lucie	RU	Fédération de Russie		
DE	Allemagne	LI	Liechtenstein	SD	Soudan		
DK	Danemark	LK	Sri Lanka	SE	Suède		
EE	Estonie	LR	Libéria	SG	Singapour		

## PROCEDE DE SEGMENTATION SPATIALE D'UNE IMAGE EN OBJETS VISUELS ET APPLICATION

L'invention concerne le domaine de l'analyse de l'information  
5 contenue dans les images couleur, provenant notamment de documents  
multimédias, et en particulier de vidéos. Cette analyse est destinée plus  
particulièrement, mais non exclusivement, à permettre une indexation de la  
documentation audiovisuelle.

Le problème consiste en particulier à réaliser un découpage  
10 temporel de séquences vidéos en scènes qui constituent des unités  
narratives de ces séquences, en vue du stockage et de la visualisation  
sélective de certaines scènes par les utilisateurs. Un tel niveau de  
compréhension n'est pas directement accessible par les méthodes de  
segmentation connues.

15 Des méthodes ont été développées pour réaliser la segmentation  
d'images de niveaux de gris ou pour la segmentation d'objets en  
mouvement dans les images d'une vidéo. Mais parmi les algorithmes de  
segmentation d'images couleur, aucun ne fournit de résultats suffisants  
lorsque les images à analyser sont tirées de la "vie réelle". Les principales  
20 méthodes peuvent être regroupées en méthodes de type pixels, de type  
contours, et de type régions.

Dans les méthodes de segmentation de type pixels, une région est  
définie comme une composante connexe d'un ensemble de pixels spécifié  
par une fonction d'appartenance, éventuellement floue, à une classe de  
25 couleur de l'Espace de Couleurs (en abrégé EdC). Ces méthodes se  
différencient principalement par la façon dont elles définissent les classes  
de couleurs et les fonctions d'appartenance à ces classes. La plus simple  
de ces méthodes est une quantification rigide de l'EdC, comme décrit dans  
les articles de C. Carson, S. Belongie, et al. "Region Based Image  
30 Querying", Proc. CVPR'97, Workshop on Content-Based Access of Image

and Video Libraries, 1997, et de J.R. Smith and S.-F. Chang. "Tools and Techniques for Color Image Retrieval.", Proc. SPIE'96, Storage & Retrieval for Image and Video Databases IV, Vol. 2670, 1996. Cependant certaines couleurs proches dans l'EdC peuvent se trouver dans des classes distinctes.

Une autre méthode de type pixel est une méthode par seuillage d'histogrammes, dans laquelle les pics et les creux apparaissant dans un ou plusieurs histogrammes correspondant aux différentes dimensions de l'EdC sont recherchés. Les creux détectés sont alors utilisés comme limites entre les classes de couleurs (comme décrit par exemple dans l'article de R. Hayasaka, J. Zhao and Y. Matsushita, "Outstanding Object-Oriented Color Image Segmentation Using Fuzzy Logic". Proc SPIE'97 Multimedia Storage and Archiving Systems II, Vol. 3229, 303-314, 1997.

Les méthodes de groupement (clustering en terminologie anglaise) de l'EdC sont des extensions multidimensionnelles des techniques précédentes de seuillage et appliquent des algorithmes de classification tels que les algorithmes de recherche de plus proches voisins, (cf. l'article de R. Ferri and E. Vidal, "Color Image Segmentation and labeling through multiediting and condensing", Pattern Recognition Letters, vol. 13, No. 8, pp. 561-568, 1992), l'algorithme de moyenne K (K-mean en terminologie anglaise) ou de moyenne floue C (Fuzzy c-mean en terminologie anglaise) (cf. l'article de Y.W. Lim, S.U. Lee, "On The Color Image Segmentation Algorithm Based on the Thresholding and the Fuzzy c-Means Techniques", Pattern Recognition, Volume 23, Number 9, pp. 935-952, 1990). Ces algorithmes permettent de rechercher des clusters potentiels de couleurs dans les images. Enfin certaines méthodes recherchent l'EdC permettant une représentation optimale des images à l'aide de techniques telles que l'analyse en composantes principales ou la transformée de Karhunen-Loève, telle que celle décrite dans l'article de S.E. Umbaugh et al, "Automatic Color Segmentation Algorithm with Application to Skin Tumor

Feature Identification", IEEE Engineering in Medicine and Biology, Vol. 12, No. 3, pp. 75-82, 1993.).

Le premier inconvénient de ces méthodes est qu'une recherche robuste de pics dans les histogrammes ou de clusters dans l'EdC n'est pas aisée, en particulier dans le cas d'images peu contrastées, et peut être coûteuse en temps de calculs. De plus, ces méthodes considèrent implicitement que si deux pixels, c'est-à-dire deux points d'image homogène, que le support soit une image vidéo ou non, appartiennent à une même région de l'image, leurs couleurs respectives appartiennent à la même classe de couleurs ou sont proches dans l'EdC. Ceci ne s'applique correctement que dans des images de type "clip-arts" ou de dessins animés, mais généralement pas dans les images réelles complexes. Au moment de l'extraction des régions, ces méthodes doivent donc analyser le voisinage immédiat des pixels pour déterminer à quelle région chaque pixel doit être attaché.

Par ailleurs, dans les méthodes de type contours, les contours sont détectés et utilisés pour déterminer les limites des régions. Cependant, les méthodes d'extraction de contours ne sont pas des techniques de segmentation par elles-mêmes et doivent être combinées avec au moins une des autres méthodes. De plus les contours obtenus dans le cas d'images peu contrastées ou très texturées sont difficilement utilisables car ils ne sont généralement pas fermés.

Dans les algorithmes de type régions, une région est définie comme une ensemble de pixels connectés satisfaisant un critère d'homogénéité donné, par exemple une zone qui ne contient qu'une couleur présente dans 95% des pixels de la région. Dans les techniques dites de découpage et fusion (split-and-merge en terminologie anglaise), les régions non homogènes sont découpées en sous-régions, puis analysées de manière récursive jusqu'à ce que le critère d'homogénéité

soit satisfait pour chacune des régions obtenues. Les régions voisines ainsi obtenues sont regroupées.

Les techniques dites de croissance de régions (region-growing, en terminologie anglaise) sont des méthodes de type régions dans lesquelles un certain nombre de zones initiales, utilisées comme germes de croissance, sont d'abord recherchées. Les pixels et régions avoisinant(es) sont itérativement englobés à ces zones initiales jusqu'à ce qu'un critère d'arrêt soit satisfait, par exemple lorsque le nombre de régions obtenues est inférieur à un seuil.

Un exemple connu de cette catégorie de méthode est l'algorithme dit des "bassins topographiques" dans lequel une image est considérée comme un relief topographique, où l'altitude de chaque point peut, par exemple, être proportionnelle à son intensité lumineuse. Le fond des bassins les plus importants est percé, et le relief est plongé dans l'eau. Les bassins se remplissent progressivement, délimitant ainsi les principales régions. Cette méthode est très sensible aux bruits et coûteuse en temps de calcul. Une autre méthode de cette catégorie est la méthode basée sur les arbres couvrants récursifs de poids minimum (Recursive Shortest Spanning Trees, en anglais, RSST en abrégé). RSST (voir l'article de O.J. Morris et al, in « graph theory for image analysis: an approach based on RSST », IEEE proceedings, vol. 1333, 146-152, 1986) considère chaque pixel comme une région initiale. Les régions dont les couleurs moyennes sont les plus proches sont fusionnées de manière récursive, en favorisant la fusion des régions de petites tailles.

La plupart des méthodes précédemment citées opèrent à l'échelle du pixel. Ceci les rend particulièrement sensibles aux variations locales d'intensité et donc aux textures.

Cette sensibilité est nécessaire pour les applications de vision artificielle ou de reconnaissance de formes pour lesquelles l'extraction des

contours exacts est primordiale, mais elle est pénalisante dans le cas de recherche de larges régions sémantiquement significatives.

L'invention vise à pallier ces inconvénients en proposant une segmentation en régions de tailles telles que les régions ainsi segmentées  
5 aient encore une signification sémantique dans le contexte de l'image. Par objet sémantique ou sémantiquement significatif, il est entendu un objet correspondant au monde réel, par exemple un visage, un ciel, etc. Plusieurs objets sémantiques peuvent composer un autre objet sémantique (par exemple une chevelure, un visage et une veste constituent une  
10 personne) ci-après appelé objet sémantique composite. Les objets sémantiques composant un objet sémantique composite peuvent aussi être eux-mêmes des objets sémantiques composites (par exemple le visage est composé, entre autres, d'un nez, d'une bouche et d'yeux).

L'utilisation d'une segmentation des images en objets ayant une  
15 valeur sémantique significative est une étape clé du procédé d'analyse et de compréhension du contenu des documents multimédias, en particulier des documents vidéo.

L'invention permet de segmenter les images en objets significatifs tout en négligeant les détails.

20 L'invention vise ainsi à obtenir une segmentation robuste en présence d'images éventuellement très texturées, et insensible aux détails non significatifs qui pourraient entraîner une sur-segmentation inutile de larges régions homogènes, par exemple une corde noire sur un mur blanc.

Pour ce faire, il est proposé de qualifier de détails à écarter les  
25 régions "non signifiantes" ayant une surface inférieure à un certain pourcentage de la surface totale de l'image (par exemple 1%) et de ne pas segmenter l'image en régions pouvant recouvrir plusieurs objets de façon à éviter sa sous-segmentation, pouvant conduire à des régions dépourvues de signification sémantique.

Le procédé de segmentation selon l'invention qui s'apparente aux techniques de type région, opère dans un premier temps à l'échelle de la région, en partant de blocs initiaux de taille réduite tout en étant considérés comme homogènes, de façon à permettre la segmentation de plus larges objets.

L'invention porte donc sur un procédé de segmentation spatiale d'une image en objets visuels, caractérisé en ce que, pour obtenir des objets ayant une signification sémantique, il comporte les étapes suivantes :

- dans une première phase, une partition de l'image en régions selon un pavage prédéterminé, une fusion de régions voisines dont la similarité, selon une première fonction de similarité, est inférieure à un premier seuil, et l'obtention de régions élargies,
- dans une deuxième phase, une fusion d'une région élargie de taille inférieure à un deuxième seuil avec une région voisine qui lui est la plus similaire selon une deuxième fonction de similarité.

Avantageusement, le procédé selon l'invention comporte également une troisième phase de fusion des régions obtenues à l'issue de la deuxième phase et qui sont similaires selon une troisième fonction de similarité.

Dans les différentes phases du procédé de segmentation selon l'invention, les fonctions de similarité utilisées sont différentes dans au moins deux des phases.

Plus précisément, l'invention a pour objet un procédé de segmentation spatiale d'une image en objets visuels qui, pour obtenir des objets ayant une signification sémantique, comporte les étapes suivantes :

- dans une première phase, une partition de l'image en régions initiales selon un pavage prédéterminé, et une fusion des régions voisines dont la similarité est inférieure à un premier seuil, selon une fonction de similarité connue entre régions ;

- dans une deuxième phase, une fusion des régions de taille inférieure à un deuxième seuil avec la région voisine la plus similaire ; et
- dans une troisième phase, une représentation de chacune des régions obtenues par un nuage de points dans un espace de représentation formé d'au moins une dimension de base caractérisant un signal électromagnétique provenant de cette région et d'une dimension caractérisant les pixels correspondant aux valeurs considérées dans les autres dimensions, avec
- un ajustement polynomial d'une courbe avec les points représentant chaque pixel de la région dans l'espace de représentation du signal électromagnétique ;
- une représentation de chaque ajustement polynomial par une suite de valeurs ;
- l'application d'une fonction de similarité entre les deux suites de valeurs ainsi définies, et
- la fusion des régions voisines dont la proximité entre les ajustements polynomiaux est inférieure à un troisième seuil.

Dans le procédé de l'invention, les régions voisines fusionnables dans chacune des phases peuvent être des régions initiales, des régions issues d'une fusion de régions initiales, ou des régions élargies issues de fusions précédentes ; les différents types de régions ainsi définies sont fusionnables entre elles à partir du moment où on leur applique la même fonction de similarité, et jusqu'à ce que toutes les similarités selon la fonction appliquée soient supérieures au seuil choisi.

Quand, dans la phase suivante, une nouvelle fonction de similarité est appliquée, la fusion des régions obtenues dans une phase précédente par application d'une fonction de similarité à seuil moins élevé est alors possible jusqu'à ce qu'aucune fusion ne soit plus possible.

Le processus peut être ainsi réitéré par application d'une nouvelle fonction de similarité à seuil plus élevé que la précédente. Les régions

fusionnables restent toujours celles qui n'ont pas fusionné au seuil moins élevé de la fonction précédente, et ce y compris des régions initiales.

De plus, la fonction de similarité entre deux régions peut être définie par la position des centroïdes des deux régions et des extrémités  
5 des deux segments de courbe représentant ces deux régions.

De la même façon, le signal électromagnétique apparaissant dans au moins deux images peut être transformé pour en extraire au moins deux composantes de mouvement entre les deux images pour le représenter, telles que valeur scalaire et orientation, les autres étapes du procédé  
10 s'appliquant à cette représentation.

Dans le procédé selon l'invention, les points d'un nuage décrivant une région sont distribués dans un espace dont trois dimensions de base sont trois combinaisons linéaires ou non linéaires distinctes des trois couleurs primaires de la synthèse additive, et une autre dimension dans  
15 cet espace étant l'effectif des pixels selon cette distribution.

Dans un autre mode de réalisation, les points d'un nuage décrivant une région sont distribués dans un espace dont trois dimensions de base sont la teinte, la saturation et l'intensité de la couleur et une autre dimension dans cet espace étant l'effectif des pixels selon cette  
20 distribution.

Dans ce cas, le calcul de similarité entre régions est réalisé :

- en utilisant l'espace teinte/saturation/intensité si une saturation moyenne est supérieure à un quatrième seuil prédéterminé ;
- dans l'espace des intensités, si la saturation est inférieure ou  
25 égale à ce quatrième seuil.

Selon un autre aspect de l'invention, une seconde segmentation à une échelle plus fine, qui peut être celle du pixel est réalisée de façon à obtenir les contours précis des limites des objets, ainsi que leur structure

interne, puis une fusion des deux segmentations est réalisée afin d'obtenir à la fois des objets sémantiquement significatifs et des contours précis pour ces objets.

Par une mise en oeuvre du même algorithme à deux échelles de  
5 résolution différentes, les résultats des deux segmentations des deux premières étapes sont cohérents et peuvent se combiner.

Ainsi l'invention concerne également un procédé de segmentation fine d'images en objets sémantiquement significatifs, consistant :

- dans un premier temps, à extraire de l'image les régions  
10 homogènes en couleur et en texture correspondant aux objets présents, en négligeant les objets de petites tailles inférieures à un seuil défini en pourcentage de la surface de l'image, et en utilisant une représentation des régions par ajustement polynomial de leur distribution de couleurs dans un espace des couleurs (EdC), à la fois lors de la segmentation et en  
15 représentation finale des objets segmentés,
- à répéter l'extraction avec une résolution plus fine, afin de préciser les contours, les régions obtenues correspondant alors à des parties d'objets,
- dans une troisième étape, une superposition des deux  
20 segmentations précédentes fournit des régions correspondant aux objets de l'image de la première étape avec les contours précis de la deuxième étape, ainsi qu'une structure interne représentative de ces objets.

Par structure interne, on entend que l'image est représentée comme une arborescence d'objets, chaque objet de niveau supérieur  
25 pouvant inclure un ou plusieurs objets de niveau inférieur.

Les deux dernières étapes peuvent naturellement être réalisées à plusieurs résolutions consécutives de façon à obtenir une description hiérarchique de la structure des objets sémantiques composites.

Préférentiellement, le calcul de similarité entre régions peut être  
30 réalisé :

- selon le procédé de l'invention si la saturation, considérée comme critère de superposition, est supérieure à un quatrième seuil prédéterminé ;

- dans l'espace teinte-saturation-intensité, si la saturation est inférieure ou égale à ce seuil ;

les seuils prédéterminés peuvent être choisis pour maintenir le nombre de régions dans un intervalle dans lequel la sur-segmentation et la sous-segmentation n'apparaissent pas et pour maintenir la distribution des tailles des régions dans un intervalle défini pour éviter la sur et la sous-segmentation.

Les paramètres de seuil, de degré de polynôme et de fonction de similarité peuvent être choisis de façon adaptative par une méthode d'apprentissage prédéterminée, en fonction de seuils de sur- et de sous-segmentation à éviter, et d'un calcul d'évaluation prédéterminé de celles-ci.

De manière générale, la fonction de similarité appliquée à l'issue quelconque des étapes du procédé inclut le paramètre de seuil au-delà duquel la fusion n'est pas réalisée.

Avantageusement, des utilisations composites du procédé selon l'invention sont mises en oeuvre :

- avec la reconnaissance d'un locuteur, pour caractériser le contenu des plans et des scènes d'une vidéo ;

- avec un alignement automatique entre le script et la vidéo, pour caractériser le contenu des plans et des scènes d'un document vidéo ou pour apprendre de façon automatique les caractéristiques des objets présents dans la vidéo et dans le script.

D'autres caractéristiques et utilisations de l'invention ressortiront de la description qui suit de modes de réalisation détaillés, accompagnés des figures annexés qui représentent respectivement :

- la figure 1, un exemple d'image obtenue à l'issue de l'application des deux premières phases de segmentation large du procédé selon l'invention ;

5 - les figures 2a et 2b, une représentation des régions d'une image dans l'espace RVB (Rouge, Vert, Bleu) par des nuages de points et des courbes d'ajustement ;

- la figure 3, les projections d'un nuage de points de l'espace RVB sur le plan (RB) et le plan (RV) pour déterminer les extrémités de l'intervalle d'ajustement ;

10 - la figure 4, la distribution des distances euclidiennes des points à leurs droites respectives pour des keyframes (images caractéristiques) et des images fixes ;

15 - les figures 5a et 5b, 6a et 6b, 7a et 7b, la segmentation en objets sémantiques sur trois images (5), (6), (7), respectivement à la fin de la deuxième phase du procédé (figures 5a, 6a et 7a) et à la fin de la troisième phase (respectivement figures 5b, 6b et 7b) ;

- les figures 8a et 8b, respectivement une segmentation dite large, obtenue avec une résolution de 16x16 pixels, et une résolution dite fine, de 4x4 pixels, sur une même image ; et

20 - la figure 9, une superposition des segmentations large et fine précédemment obtenues.

25 L'invention est d'abord décrite comme un algorithme de la segmentation de larges régions. La segmentation fine est un mode plus particulier pouvant être réalisé par le même algorithme. Un algorithme utilisant deux résolutions, et combinant les deux segmentations obtenues, est décrit plus loin.

Selon un premier exemple de réalisation, l'image est d'abord découpée en une grille de blocs dits initiaux, de taille adaptée, égale à 0,25% de la surface de l'image courante, c'est-à-dire :

- suffisamment petits pour que les objets de taille supérieure à un seuil (1% dans l'exemple de réalisation), en dessous duquel ils sont considérés comme des détails, soient segmentés, et

- suffisamment larges pour limiter les temps de calcul, qui augmentent de manière « surlinéaire » avec le nombre de blocs initiaux.

A partir de ce découpage, un histogramme des couleurs de chaque bloc initial est calculé dans l'espace RVB (Rouge Vert Bleu). Ce calcul consiste à comptabiliser le nombre de pixels ayant chacune des couleurs (des 256 couleurs dans l'exemple de réalisation) apparaissant dans l'image, chaque couleur étant la valeur prise par l'une des coordonnées de l'espace RVB.

Dans la première phase de fusion, chaque région est évaluée par un calcul de distance entre son histogramme et ceux des régions adjacentes. Les calculs des distances sont réalisés selon la norme d'ordre 1 ( $L_1$ ), égale à la somme des valeurs absolues des différences au 1er degré des valeurs des histogrammes prises deux par deux, les normes d'ordre  $n$ , ( $L_n$ ), étant celles de Minkowski (égales à la puissance  $1/n$  de la somme des valeurs absolues des mêmes différences à la puissance  $n$ ).

La région courante est fusionnée avec celle dont l'histogramme est le plus proche du sien, mais uniquement si la distance qui sépare leurs histogrammes est inférieure à un seuil élevé qui, dans cet exemple de réalisation, est fixé à 50% de la distance maximale possible de cet exemple de réalisation. Plus le seuil est élevé, plus le niveau de similarité des régions fusionnées est élevé. La fusion est réitérée jusqu'à ce que toutes les distances entre régions adjacentes soient supérieures à ce seuil.

A la fin de cette première phase, les régions restantes sont soit des petites régions, c'est-à-dire correspondant à des détails (taille inférieure à 1% dans l'exemple de réalisation), soit des régions plus larges et homogènes, du fait du seuil de fusion élevé (50% dans l'exemple de réalisation). Ces larges régions peuvent éventuellement être très

texturées, c'est-à-dire avec une granularité de texture de l'ordre de la taille des blocs initiaux.

La première phase est suivie d'une deuxième phase de fusion des seules petites régions, de taille inférieure à 1% dans l'exemple de réalisation. Ces "détails" sont fusionnés automatiquement avec leur plus  
5 proche voisine par suppression du seuil de fusion de 50 % utilisé lors de la première phase, de sorte que tous les détails sont intégrés à leur région englobante ou à leur plus proche voisine.

A ce point de l'algorithme, un graphe de régions larges et  
10 homogènes est obtenu, correspondant à de larges parties d'objets situés dans l'image, et contenant éventuellement les détails intégrés lors de la deuxième phase. La figure 1 montre le résultat de la segmentation à la fin de la deuxième phase. Le tumulus représenté T a été fusionné malgré la granularité importante de la texture de cette région grâce à l'utilisation  
15 d'une taille de blocs initiaux (1 %) de l'ordre de grandeur du grain de la texture.

L'entrée E du tumulus étant plus petite que le seuil de 1% dans l'exemple de réalisation, elle a été segmentée lors de la première phase, puis a été fusionnée lors de la deuxième phase puisque, évaluée comme  
20 détail, elle ne pouvait pas être considérée comme un objet sémantique significatif par le présent algorithme.

Ainsi, en partant de blocs initiaux de taille suffisamment importante, des comparaisons de distribution de couleurs et non des comparaisons de couleurs sont réalisées, ce qui rend l'algorithme  
25 insensible aux textures de granularité de l'ordre de celle des blocs initiaux.

En revanche, le ciel C qui est composé d'un dégradé progressif de plus de 50 variations de bleus est dit sur-segmenté car, étant a priori un objet sémantiquement unique, on devait s'attendre à une segmentation unique pour ne former qu'une seule région.

Des objets, comme le ciel de cette figure 1, peuvent être sur-segmentés, les régions qui les composent restant « perceptuellement » similaires. La troisième phase permet de dépasser ce stade.

La sur-segmentation du ciel C illustre les limites de l'utilisation des  
5 histogrammes de couleurs : ils sont sensibles aux effets optiques tels que l'illumination, les variations d'éclairage ou les dégradés, comme décrit dans l'article de M. Stricker and M. Orengo, "Similarity of Color Images", Proc. SPIE'95, Storage & Retrieval for Image and Video Databases III, 1995). De plus, comme les régions sont homogènes, leurs histogrammes  
10 correspondants sont "creux" (c'est-à-dire présentent une proportion importante de valeurs proches ou égales à 0) et ne peuvent donc être efficacement comparés à l'aide des mesures de distances telles que L1 ou L2.

Une autre représentation des régions est donc réalisée dans une  
15 troisième phase, afin de permettre non seulement de comparer les distributions de couleurs, mais aussi les couleurs dans l'EdC (Espace des Couleurs).

Une représentation des régions d'une image dans l'espace RVB (Rouge, Vert, Bleu) est illustrée aux figures 2a et 2b.

20 Sur la figure 2a, chaque région est représentée par la moyenne statistique des valeurs de son histogramme formant sa couleur moyenne, correspondant au centroïde C de chaque nuage de points N1 à N6, correspondant aux distributions de couleurs des régions, et par un intervalle polynomial, P, qui donne une estimation des variations de  
25 couleurs au sein de la région.

Les distributions des couleurs qui présentent une forte corrélation ont une forme curviligne, ce qui justifie l'estimation de ces régions par un ajustement polynomial, avec un intervalle de domaine de définition du polynôme. La figure 2b représente les nuages de points correspondants à  
30 la distribution des couleurs apparaissant dans l'image composée de 3

régions sémantiques principales R1 à R3, ainsi qu'une courbe C1 (en pointillés) correspondant à l'estimation polynomiale de la région R1.

A la fin de la deuxième phase, les nuages des régions obtenus sont des sous-ensembles, ou sous-régions, des régions finales R1 à R3.

5 Dans le cas présent, 6 sous-régions N1 à N6 ont été extraites. Un ajustement polynomial est calculé pour les régions finales, ainsi qu'un intervalle sur leur domaine de définition. Dans le cas idéal, le polynôme P d'une région finale et ceux  $P_i$  des sous-régions (obtenues en fin de deuxième phase) qui la composent, sont les mêmes. Et l'intervalle de P est  
10 l'union des intervalles des  $P_i$ .

La troisième phase de l'algorithme fusionne les régions dont les ajustement polynomiaux sont similaires, et ayant des intervalles sur le domaine de définition similaires, consécutifs, ou se chevauchant.

En pratique, pour limiter le nombre de calculs à réaliser, une droite  
15 est approximée en utilisant la méthode classique de régression linéaire. Dans le présent exemple de réalisation de l'algorithme, les nuages de points des régions obtenues (Figure 2b) à la fin de la deuxième phase sont alors représenté par un segment de droite obtenu par un ajustement linéaire et par le centroïde du nuage N1 à N6 correspondant, qui n'est pas  
20 nécessairement le centre du segment.

Chacune des régions est représentée par sa couleur moyenne et par un segment S1 à S6 de droite porté par la droite de régression linéaire du nuage de points correspondant (figure 2b).

Pour chaque région, tel qu'illustré sur la figure 3, le nuage de  
25 points de coordonnées  $(r, v, b)$  de l'espace RVB est projeté en  $N(r, b)$  sur le plan (RB) ainsi qu'en  $N(r, v)$  sur le plan (RV). Dans ces plans, nous calculons respectivement les droites de régression linéaire de b en r, notée D1 sur la figure, et de v en r, notée D2. Les deux droites obtenues sont les projections sur les deux plans (RB) et (RV) de la droite cherchée, notée  
30 D3.

Les extrémités E1 et E2 du segment de droite représentatif S sont les projections sur la droite de régression D3 des points extrêmes du nuage de points N. Dans ces conditions, le segment ne va pas au-delà de la projection des points les plus extrêmes du nuage de points.

5 Pour éviter que la longueur du segment ne soit influencée de façon importante par la présence de couleurs provenant par exemple de détails, la distance euclidienne de ces extrémités au centroïde du nuage est de préférence limitée à un seuil dans l'espace RVB, seuil égal à 1,5 fois l'écart type de la distribution des couleurs pour la région considérée dans  
10 l'exemple de réalisation.

Une telle estimation des distributions des couleurs des régions dans l'EdC par un ajustement linéaire a été testée en mesurant les coefficients de régression linéaire obtenus lors des ajustements, ainsi que la longueur des segments obtenus et la distance euclidienne moyenne des  
15 points à leurs droites respectives. Ces mesures sont réalisées sur une collection de 4000 images caractéristiques (keyframes en terminologie anglaise) extraites de 4 films, et une collection d'environ 1000 images fixes obtenues sur les sites du réseau internet.

Les coefficients de régression moyens obtenus sont égaux à 0,87  
20 pour les keyframes et à 0,84 pour les images fixes.

La distribution des distances euclidiennes des points à leurs droites respectives est représentée sur la figure 4. Pour les keyframes K, respectivement les images fixes F, la distance moyenne est de 3,17, respectivement 4,88, pour une longueur moyenne de segment de 74,63,  
25 respectivement 90,5.

Ces résultats montrent une forte corrélation linéaire entre les couleurs présentes dans les régions, qui justifie la représentation par approximation polynomiale adoptée dans le présent algorithme.

La fusion des régions lors de la troisième phase de l'algorithme est réalisée en comparant non plus les histogrammes de couleurs, mais les segments représentatifs obtenus à la deuxième étape.

La comparaison des segments est réalisée dans l'espace Teinte-Saturation-Intensité HSI (initiales de « Hue-Saturation-Intensity » en terminologie anglaise). Cet espace est perceptuellement uniforme, car fournissant une représentation linéaire des variations de fréquence spectrale d'une couleur, alors que l'espace RVB, ne fournissant pas une telle représentation, n'est pas adapté à une telle comparaison.

Cette comparaison exploite le fait que les variations de Teinte et de Saturation dans les régions sont faibles, alors que les variations d'intensité sont importantes. Les variations moyennes observées sur toutes les régions extraites de toutes les images des collections précédentes sont présentées dans le tableau 1. Les variations de teintes sont exprimées en degrés, les autres paramètres en distance euclidienne dans l'espace RVB.

	Teinte	Saturation	Intensité
Keyframes	15.17	9.56	41.76
Images	20.97	18.35	48.48

-- TABLEAU 1 --

Pour limiter, lors de la troisième étape, la fusion aux régions perceptuellement similaires, et donc appartenant potentiellement aux mêmes objets de la scène, seules les régions dont les différences de Teinte et de Saturation entre les centroïdes sont inférieures à un seuil donné sont fusionnées.

Dans l'exemple de réalisation, la différence maximale de teinte est fixée à un seuil égal à 7,5°, et la différence maximale de saturation à un seuil de 15%. Comme représenté sur l'exemple de la figure 3, les régions voisines R1, R2 et R3, satisfaisant ces critères sont fusionnées. Ces

segments représentatifs sont proches du meilleur ajustement polynomial, pouvant être obtenu dans le cas d'un ajustement d'ordre supérieur à un.

Lorsque la saturation moyenne est faible, par exemple inférieure à environ 7% de la saturation maximale, la notion de teinte n'est plus représentative et les segments sont alors uniquement comparés sur la base de leur intensité moyenne et de leur variation en intensité. Ce cas correspond aux couleurs "proches" des gris. Ainsi, la méthode de comparaison consiste principalement à comparer les niveaux de gris moyens (i.e. l'intensité moyenne) et les variations en intensité (i.e. les textures).

Cette méthode de comparaison est une approximation d'une mesure de similarité entre ajustements polynomiaux, mais est néanmoins valide puisque le coefficient de corrélation linéaire au sein des nuages de points correspondant aux régions est élevé.

Les figures suivantes présentent la segmentation  $\Sigma$  en objets sémantiques sur trois images (5), (6), (7), respectivement à la fin de la deuxième phase du procédé (figures 5a, 6a et 7a) et à la fin de la troisième phase (respectivement figures 5b, 6b et 7b).

Il apparaît que la fusion des régions similaires, obtenue lors de la troisième étape, fournit une segmentation réaliste des images (5), (6), (7).

Un autre mode de réalisation concerne la segmentation fine d'objets obtenus à l'aide du procédé de fusion précédent, visant à obtenir le contour fin de ces objets ainsi que leur structure interne. Le même algorithme est utilisé avec au moins deux résolutions différentes, l'une dite fine et l'autre dite large. Cette résolution plus fine est obtenue uniquement en utilisant des blocs initiaux de taille plus petite, par exemple 4x4 pixels. L'algorithme réalise alors la segmentation en découpant à l'échelle du pixel, ce qui la rend plus sensible aux contours et aux textures.

Contrairement aux méthodes de l'état de la technique évoquées plus haut, même lors de la segmentation fine, les zones dont la surface

représente moins d'une valeur seuil, fixée à 1% dans l'exemple de réalisation dans l'exemple de réalisation, de la surface de l'image ne sont pas prises en compte. Cela permet en particulier d'extraire les petits objets qui n'auraient pas été détectés par la première segmentation du fait de la  
5 résolution plus importante. Les figures 8a et 8b présentent respectivement une segmentation dite large  $\Sigma$ , obtenue avec une résolution de 16x16 pixels, et une résolution dite fine  $\Sigma_f$  de 4x4 pixels, sur une même image.

Puis une superposition permet d'obtenir une segmentation finale optimisée  $\Sigma_o$ , telle que représentée en figure 9 ; les deux segmentations  
10 sont obtenues à l'aide du même algorithme avec, comme paramètre d'entrée relatif à la résolution, 4 X 4 pixels et 16 X 16 pixels.

La superposition conserve les contours 10 des régions fines incluses dans les régions ayant une signification sémantique. Cependant, les régions fines peuvent ne pas être systématiquement incluses dans la  
15 région large correspondante, car pouvant résulter de détails obtenus grâce à la résolution plus fine, par exemple les barres 11 sur le mur à l'arrière du personnage de la figure 9.

Pour plus de clarté, nous utilisons le terme de région pour désigner les larges régions sémantiquement significatives obtenues par une  
20 segmentation avec de larges blocs initiaux, et le terme de zones pour désigner les régions fines obtenues avec de petits blocs. Le résultat final (figure 9) est obtenu par la superposition des deux résultats intermédiaires. Dans le cas idéal, chaque région serait composée exactement (au contour près) de l'ensemble des zones qu'elle contient spatialement.

25 Trois cas peuvent cependant se présenter :

- Une zone est spatialement incluse dans une région et sa représentation est proche (au sens de la mesure de similarité utilisée lors de la troisième phase de l'algorithme de base) de celle de cette région. Dans ce cas, on considère que la zone fait partie de la région.

• Du fait de la résolution plus précise de la segmentation fine, une zone peut correspondre à un détail de l'image n'ayant pas été extrait par la segmentation grossière (ex : les barres 11 sur le mur à l'arrière du personnage de la figure 9). Dans ce cas la distance entre les représentations de la zone et de la région est élevée, et on considère que la zone ne fait pas partie de la région, mais forme une région par elle-même. Ces zones peuvent éventuellement être utilisées pour l'étude de la structure spatiale des régions.

• Une zone n'est pas majoritairement incluse dans une région (en pratique, un seuil dépendant de la taille de la zone est fixé) mais s'étend sur plusieurs régions (ex : le col 12 de la chemise du personnage de la figure 9). Dans ce cas, la zone fait partie de la région la plus similaire, ou est considérée comme région à part entière si aucune des régions avoisinantes n'est assez similaire.

Des exemples d'application du procédé selon l'invention sont décrits ci-après.

#### Exemple 1 : Caractérisation des objets; classification.

L'ensemble des caractéristiques des objets segmentés est conservé en vue d'une analyse du document et/ou d'une indexation des images pour permettre la recherche de plans sur leur contenu en terme d'objets sémantiques et d'actions de ces objets. L'ensemble des caractéristiques de chaque région (couleur, texture, taille, position, indice de forme, mouvement,...) est très compact. Dans le cas d'un ajustement linéaire, un résumé des caractéristiques précédentes peut être stockés dans moins de 20 octets (La position de la couleur moyenne et du segment représentatif en nécessitant 9). Des représentations plus complètes peuvent nécessiter quelques dizaines d'octets par objet.

Une image peut être représentée sous une forme résumée par une liste de descripteurs des principaux objets qu'elle contient, chaque descripteur incluant notamment la position et le mouvement de l'objet. Les

critères de choix des principaux objets peuvent être par exemple la taille, la différence de couleur avec les objets voisins, le mouvement, ou des connaissances sémantiques du type "l'objet X est important". Sur la base d'une représentation de 4 objets dans une image, par exemple, la  
5 représentation de l'image peut être aussi compacte que 80 octets. Pour caractériser un objet, il est utile de connaître non seulement le descripteur de cet objet, mais également les descripteurs des objets voisins, puisqu'un objets peut aussi être caractérisé par son contexte (ex : un avion dans le ciel.).

10 La sémantique de certains des objets segmentés par la méthode proposée peut être aisément extraite à l'aide de leurs caractéristiques visuelles dans un certain nombre de cas simples (ex : détection de ciels de jours, d'éclairages, de peau, ...). Cependant, l'association d'une sémantique aux objets peut également s'appuyer sur l'apport de  
15 connaissances externes. Par exemple : "un ciel est un objet bleu ou gris peu texturé, en général en haut d'une image".

Le problème de la caractérisation d'un objet sémantique est un problème connu de classification ou de groupement (clustering) de points dans un espace multi-dimensionnel. Cette classification peut s'effectuer  
20 avec ou sans apprentissage, en mode supervisé ou non. Dans le procédé du présent brevet, cette classification s'appuie sur une représentation compacte de l'objet et le cas échéant des objets environnants, dans un espace multi-dimensionnel.

Les méthodes connues de classification qui peuvent être utilisées  
25 sont les méthodes classiques d'analyse des données, les méthodes neuronales et les méthodes par algorithmes génétiques.

Dans certaines méthodes d'analyse des données, les nuages de points voisins sont caractérisés comme des clusters et projetés dans un espace de représentation plus petit adéquat. La caractérisation des objets  
30 du cluster peut alors se faire à partir de la description par l'utilisateur d'un

ou plusieurs objets du cluster. Le système d'indexation généralise la caractérisation d'un ou plusieurs objets décrits par des points du cluster, à des objets décrits par d'autres points de ce cluster.

Dans l'approche neuronale, le système apprend un « classifieur »,  
5 qui permet de diviser l'espace de représentation en clusters, à partir d'un ensemble d'exemples. Les exemples peuvent être fournis par un des utilisateurs du système pendant l'apprentissage ou en cours d'utilisation.

Dans l'approche génétique, il existe un ensemble de classifieurs initiaux, caractérisés chacun par une fonction de similarité prise dans un  
10 ensemble de fonctions possibles, et par des seuils. Ces classifieurs sont représentés par une signature qui est une chaîne de bits. Les classifieurs initiaux peuvent être tirés au hasard ou fournis par des utilisateurs. L'utilisateur ou le système détermine quels classifieurs ont donné une réponse appropriée. Les classifieurs qui ont participé aux bonnes  
15 réponses sont hybridés par recombinaison de la signature de deux de ces classifieurs. Des modifications aléatoires des signatures ou "mutations", peuvent également être appliquées lors de la création de nouveaux classifieurs. Pour certaines classes d'applications, ce processus converge vers une population de classifieurs proche de l'optimum.

20 La quatrième méthode de classification envisageable à partir de la segmentation de l'image en objets sémantiques, est la recherche d'objets visuels similaires à un ensemble d'exemples donnés, à partir d'une similarité des caractéristiques de couleurs, de formes, etc. La requête initiale obtient une fonction globale de similarité, en calculant une somme  
25 de fonctions de similarité appliquées indépendamment à différents critères, chacune étant pondérée par une valeur appelée poids. Cette requête initiale peut être enrichie de façon connue en permettant à l'utilisateur de spécifier quels sont les réponses satisfaisantes ou non. Une technique générique d'enrichissement d'une requête vectorielle à partir de ces  
30 réponses est connue. Dans certaines variantes de cette technique, elle

peut s'appuyer sur l'estimation des modifications souhaitables des poids des différentes fonctions de similarité, par des méthodes dérivées des probabilités bayésiennes. La recherche vectorielle par similarité est proposée par plusieurs projet d'indexation de la vidéo, mais ces projets ne  
5 s'appuient pas sur une segmentation spatiale des objet sémantiques et une mesure de similarité selon notre procédé.

Selon la présente approche, l'application de ces méthodes est facilitée par la taille réduite du descripteur, et par la possibilité pour l'utilisateur d'indiquer des exemples et des contre-exemples, et le cas  
10 échéant d'indiquer si la réponse est satisfaisante ou non.

Un script décrivant le contenu de chaque plan d'une vidéo est aligné sur les limites de plans de la vidéo par des méthodes connues. Ce script décrit chaque objet d'un plan et ses actions. En corrélant la présence d'un objet dans le script et dans la vidéo, il est possible de déterminer avec  
15 une certaine probabilité quel objet de la vidéo correspond à un objet du script et quelles sont ses actions. A partir de cette information, on dispose d'exemples de ce type d'objet, qui permettent de construire automatiquement un classifieur pour cet objet.

Une autre utilisation de ces méthodes dans notre procédé est d'utiliser la segmentation pour annoter les objets par des caractéristiques  
20 objectives ou subjectives. Pour reconnaître la présence d'une de ces caractéristiques dans un objet ou une partie d'une vidéo, il est possible de choisir de façon automatique comme exemples les objets visuels qui sont annotés par cette caractéristique, et de procéder ensuite à l'apprentissage  
25 d'un des classifieurs précédents.

Dans le cas particulier où l'objet correspond à un personnage, la reconnaissance du locuteur par des méthodes connues d'analyse audio permet de choisir comme exemple plusieurs instances d'un même objet, et de procéder ensuite à l'apprentissage des caractéristiques de cet objets  
30 selon l'une des méthodes mentionnées précédemment.

Dans tous les cas, la segmentation et la classification des objets d'une scène et de leur mouvement permet de générer un script de la scène qui décrit les objets, les actions, et l'ambiance de la scène.

Exemple 2 : Segmentation temporelle de vidéo en séquences

5 De façon connue, une vidéo est le plus souvent structurée en plans, séparés par des coupures (cuts en anglais) ou par des effets spéciaux (fondus, volets). Un plan est une suite continue d'images effectuée en une seule prise par une seule caméra. La segmentation d'une vidéo en plans est utile notamment pour naviguer dans la vidéo à partir d'un  
10 interface appelé "story-board", qui représente chaque plan par une image caractéristique.

Cependant, plusieurs plans peuvent avoir un contenu sémantique voisin, et d'autre part le nombre de plans dans une vidéo est souvent très élevé, ce qui est défavorable à une navigation efficace dans la vidéo à  
15 partir du story board.

Plusieurs auteurs ont proposé de représenter la vidéo en vue d'une description de son contenu et de la navigation dans ce contenu, comme une suite de séquences. Selon cette approche, une séquence est une suite de plans décrivant un même environnement et les mêmes  
20 personnages. La séquence est une unité sémantique appropriée pour la description de contenu et la navigation dans la vidéo.

Des méthodes pour la segmentation automatique d'une vidéo en séquences ont déjà été proposées. Par exemple, un changement dans l'évolution de la durée des plans peut caractériser un changement de  
25 séquences.

Une autre méthode de segmentation en séquences est basée sur la détection d'objets caractéristiques. Par exemple, un changement de séquence est souvent lié à un changement d'environnement, par exemple

entre intérieur / extérieur / jour / nuit. La détection d'un objet de type ciel de jour ou ciel de nuit ou éclairage permet éventuellement de caractériser un plan comme tourné en extérieur jour ou extérieur nuit.

Dans le procédé selon la présente invention, la segmentation en  
5 objets sémantiques puis la caractérisation d'un certain nombre d'objets par les méthodes du paragraphe précédent permet de détecter des limites de séquences.

Deux autres types d'intervalles temporels sont les groupes de plans (prises de vue, groupements) et les sujets. Les groupes de plans ont  
10 les mêmes propriétés que les séquences, mais ne sont pas formés nécessairement de plans contigus. Les sujets sont une suite de séquences portant sur un même thème. La détection de sujets est particulièrement intéressante pour caractériser des intervalles temporels dans des vidéos documentaires ou d'actualités.

15 La segmentation en sujets selon la présente application s'appuie sur la segmentation en séquences selon l'approche décrite précédemment. La détection d'une limite de sujet se fait à partir d'une ou plusieurs des méthodes suivantes :

- Une méthode syntaxique connue
- 20 • La présence d'un objet sémantique appartenant à une classe d'objets caractérisant un certain sujet
- L'extraction du texte des dialogues et des commentaires par des méthodes connues de dictée vocale, les mots de ces textes étant ensuite classés par concepts à l'aide d'une méthode telle que celle publiée  
25 auparavant par Dumais et Foltz (Comm. ACM ).

Une fois la vidéo segmentée en sujets, la caractérisation du sujet par un ou plusieurs descripteurs s'appuie sur les mêmes méthodes.

#### Exemple 3 : Segmentation d'objets sémantiques composites

30 Les objets sémantiques composites sont composés d'une ou plusieurs régions sémantiques segmentées par la méthode décrite ci-

dessus. Dans la figure 8, le personnage est un objet sémantique composite composé du casque, du visage, de la veste, du col de chemise, ... Il est intéressant de pouvoir le retrouver par une région quelconque, par exemple par le casque, ou de visualiser et d'annoter le personnage complet et pas seulement le visage.

La segmentation de tels objets ne peut en général être réalisée uniquement à partir d'une image fixe.

Il est dans certain cas possible, par exemple lorsqu'une focale longue a été utilisée pour la prise de vue, de séparer les objets de l'avant plan qui sont nets, des objets de l'arrière plan qui sont susceptible d'être flous. Ce type de segmentation n'est cependant pas utilisable systématiquement, et est susceptible de conduire à des sous-segmentations importantes (exemple : l'arrière plan segmenté comme un seul objet).

La segmentation des objets en mouvements peut aussi être utilisée pour la segmentation des objets composites dans les vidéos. Les méthodes de segmentation basées sur le mouvement et de suivi d'objets ont récemment fait l'objet de nombreuses publications ( voir Proceedings of the Workshop on Image Analysis for multimedia Interactiv Services – WIAMIS'97, Louvain-la-Neuve, Belgique, juin 1997). Une limitation des méthodes basées principalement sur le mouvement, par rapport à notre approche basée sur les objets sémantiques, est que la méthode basée sur le mouvement n'est pas toujours applicable (objets statiques) et qu'elle ne fournit pas d'informations de structure.

Selon la présente approche, il est possible d'utiliser une représentation sur plusieurs dimensions du mouvement et de la couleur (par exemple : deux dimensions pour le mouvement, une pour le nombre de blocs pour chaque valeur des deux dimensions précédentes). Cependant dans une autre mise en œuvre utile de notre procédé, la segmentation est faite uniquement sur la couleur, la distribution du

mouvement est calculée sur chaque objet ou partie d'objet. Cette distribution (par exemple la moyenne et l'écart type des vecteurs de mouvement) est utilisée pour définir des objets composites à partir d'un même mouvement des différentes parties.

5 Les différences entre les caractéristiques de mouvement dans différentes parties d'un objet sémantique peuvent aussi être utilisées pour décrire un mouvement complexe ou une action de cet objet. Par exemple, le mouvement d'un bras dans un personnage n'est pas forcément le mouvement moyen de l'objet.

10 Après compensation des mouvements de caméra (travellings, zooms, ...) les informations de mouvements peuvent être utilisées de 2 façons :

- la segmentation des objets en mouvement retourne une segmentation en région qui est nécessairement un sur-ensemble de la segmentation obtenue sur les images fixes par notre algorithme. Les régions appartenant à un même objet en mouvement sont regroupées pour former un objet composite ;
- les régions obtenues par la méthode de segmentation présentée ci-dessus et présentant les mêmes caractéristiques de mouvement (vitesse, trajectoire) sur des images consécutives sont regroupées pour former un objet composite.

15 Si l'analyse du mouvement est importante pour l'amélioration de la segmentation en objets composites, elle n'est pas suffisante dans certains cas, notamment dans la cas de plans ou scènes relativement statiques.

25 Une troisième méthode basée sur la cooccurrence des régions dans les images est proposée : si la combinaison de régions, par exemple casque-visage-veste, apparaît régulièrement dans des suites de plans, alors ces régions peuvent être associées avec une probabilité importante de cooccurrence. Les objets étant représentés dans un espace des caractéristiques citées précédemment, la probabilité de cooccurrence des

30

objets sémantiques segmentés peut être calculée, par exemple par les méthodes statistiques connues utilisées pour l'indexation de documents textuels.

Enfin, les objets voisins peuvent être regroupés en objets  
5 composites à l'aide de leur valeur sémantique.

Exemple 4 : Suivi d'un objet à travers une ou plusieurs scènes

Les méthodes précédentes permettent de retrouver un même objet sémantique à travers plusieurs images successives d'une scène d'une vidéo, ou même à travers plusieurs scènes situées dans différents  
10 passages de cette vidéo. Cette caractérisation se fait à partir des méthodes de similarité entre objets ou régions visuelles décrites précédemment, et compte tenu des déplacements de l'objet : un objet semblable situé à la même place dans l'image suivante a plus de chance d'être le même objet que s'il est dans une partie opposée de l'image.

15 Le suivi d'objet (ou traçage d'objets) dans une vidéo est un problème connu, qui fait l'objet de plusieurs travaux d'autres auteurs. Dans notre approche, le fait de disposer d'objets sémantiques simples ou composites limite le nombre d'objets à suivre, d'autre part nous utilisons une méthode spécifique de recherche de similarité entre plusieurs  
20 occurrences d'un objet, comme nous l'avons décrit précédemment.

Dans certains cas, le suivi d'objets permet de détecter des actions. Par exemple le fait que deux objets se déplacent ensemble puis sont séparés, traduit fréquemment le fait qu'un des objets a déposé l'autre au cours de l'intervalle. Le fait qu'il s'agit d'objets sémantiques accroît la  
25 qualité de cette détection d'actions.

Exemple 5 : Sélection d'objets en vue du stockage sur le système d'un utilisateur

Lors de l'accès intelligent à une base d'images, ou une grande base de vidéos, ou un ensemble de programmes audiovisuels diffusés, il  
30 est utile de pouvoir stocker une partie de cette base ou de ces

programmes intéressante pour un utilisateur, sur un système de stockage, par exemple au site de production du programme ou chez l'utilisateur.

Dans le présent procédé, les parties de la vidéo stockées sur ce système de stockage sont des intervalles temporels ou des ensembles d'images caractérisés par la présence de descripteurs de séquences ou de sujets vérifiant une requête de l'utilisateur ou du système, ou par la présence d'objets visuels ou sonores vérifiant une telle requête.

Ces parties sont stockées dans un certain ordre, qui dépend soit de la pertinence des objets trouvés, soit de règles de transition entre objets définies par le système ou par l'utilisateur, soit un ordre aléatoire. Les règles de transition peuvent être extraites à partir des associations usuelles faites par un utilisateur, ou des transitions faites usuellement par celui-ci.

La requête utilisée a pour but de retrouver un objet (ou une séquence ou un sujet) dans laquelle on retrouve avec un degré de pertinence et de réalisation plus ou moins élevé un ensemble de caractéristiques de contenu présents dans la requête ou dans un jeu d'exemples associé à la requête. Les objets ou les segments temporels recherchés peuvent être ceux pour lesquels soit l'utilisateur a manifesté un intérêt, par exemple par la consultation d'objets semblables lors de sessions antérieures, soit un utilisateur semblable a manifesté un intérêt. Deux utilisateurs sont semblables par rapport à cette application, s'ils manifestent un intérêt pour un même ensemble de documents ou d'objets audiovisuels. Dans ces deux cas, les requêtes qui serviront à sélectionner les objets ou les segments temporels peuvent être produites automatiquement par le système.

#### Exemple 6 : Utilisation pour un système de compression et de composition d'objets vidéo

Dans un système de compression de vidéos basé sur une représentation par objets, il est utile de pouvoir décrire une image comme

un ensemble d'objets. Un objet qui intéresse davantage un utilisateur peut être transmis avec un taux de compression plus faible qu'un autre objet (tel que l'arrière plan). D'autre part, une scène vidéo peut être éditée de façon à juxtaposer plusieurs objets provenant de scènes différentes, ou à  
5 supprimer certains objets.

Pour réaliser ces objectifs, la disposition d'une segmentation en objets sémantiques est utile. Dans notre approche, les méthodes de segmentation utilisées permettent d'accéder à une structure arborescente des objets, depuis l'intervalle temporel ou l'image, puis des objets  
10 composites, jusqu'à la structure interne de ces objets, comme nous l'avons décrit précédemment.

Cette approche permet d'appliquer les méthodes d'un système de représentation de vidéo par objets, d'une façon efficace, et avec une granularité qui varie de l'objet composite à la structure fine.

15 L'invention n'est pas limitée aux exemples décrits et représentés. D'autres utilisations et applications sont ci-après définies :

- utilisation selon laquelle une région à signification sémantique est également caractérisée par des régions plus petites contenues dans la première ;
- 20 - pour étiqueter les objets dans une image par une description du type et du contenu de ces objets ;
- pour détecter les objets communs dans plusieurs plans d'une vidéo, en vue de grouper ces plans en groupes de plans à signification sémantique ;
- 25 - pour détecter les suites de plans réalisés dans un même contexte, appelés scènes ou séquences, et les groupes de plans réalisés dans un même contexte, appelés groupes de plans ou clusters ;
- l'utilisation conjointe avec la dictée vocale pour caractériser le contenu des plans et des scènes d'une vidéo ;

- application à des images fixes et animées, conjointement avec des annotations effectuées par l'utilisateur, pour caractériser les objets de la vidéo par des caractéristiques objectives et subjectives ;
- application à une base d'images, pour caractériser un objet composite pour l'association fréquente de plusieurs composantes de l'objet dans la même image ;
- l'utilisation pour décrire un objet audiovisuel comportant des images, dans un format de représentation décrivant notamment la position des objets sémantiques contenus dans l'objet audiovisuel, ces objets sémantiques étant caractérisés par un ensemble de caractéristiques sémantiques ;
- l'utilisation pour décrire un objet audiovisuel comportant des images, dans un format de représentation décrivant notamment les actions des objets sémantiques contenus dans l'objet audiovisuel ;
- l'utilisation pour sélectionner les objets d'un flot d'objets audiovisuels, devant être stockés dans le système de stockage d'un utilisateur de ce flot audiovisuel en vue de l'accès ultérieur à ces objets ;
- l'utilisation des procédés selon l'invention, dans lesquels le degré de vraisemblance ou de réalisation de la caractéristique est respectivement caractérisée ou représentée par un nombre.
- l'utilisation pour la compression et la transmission d'images fixes et animées avec un degré de compression variable selon l'interaction actuelle ou antérieure de l'utilisateur avec un objet sémantique donné ;
- l'utilisation pour la représentation d'images fixes et animées comme un ensemble d'objets sémantiques dans une scène dont la composition peut être modifiée par l'utilisateur.

## REVENDEICATIONS

1. Procédé de segmentation spatiale d'une image en objets visuels, caractérisé en ce que, pour obtenir des objets ayant une signification  
5 sémantique, il comporte les étapes suivantes :

- dans une première phase, une partition de l'image en régions selon un pavage prédéterminé, une fusion de régions voisines dont la similarité, selon une première fonction de similarité, est inférieure à un premier seuil, et l'obtention de régions élargies,

10 - dans une deuxième phase, une fusion d'une région élargie de taille inférieure à un deuxième seuil avec une région voisine qui lui est la plus similaire selon une deuxième fonction de similarité.

2. Procédé selon la revendication 1, caractérisé en ce qu'il comporte une troisième phase, de fusion des régions obtenues à l'issue de la  
15 deuxième phase et qui sont similaires selon une troisième fonction de similarité.

3. Procédé selon la revendication 1 ou 2, caractérisé en ce que les fonctions de similarité dans au moins deux des phases sont différentes.

4. Procédé selon l'une des revendications 1 à 3 dans lequel les  
20 régions voisines fusionnables dans l'une quelconque des phases peuvent être des régions initiales ou des régions issues d'une fusion selon une fonction de similarité différente.

5. Procédé selon l'une des revendications 1 à 4, caractérisé en ce que pour une fonction de similarité on retient une méthode comportant :

25 - une distribution des pixels d'une région dans un espace de représentation de manière à former un nuage de points dans cet espace, chaque point de cet espace y représentant un pixel,

- cet espace de représentation comportant au moins une dimension de base relatives à des caractéristiques physiques des pixels,

30 - une extraction d'un ensemble de grandeurs statistiques relatives à

ce nuage, et

- la définition d'un critère de ressemblance entre ces grandeurs statistiques.

5 6. Procédé selon la revendication 4, caractérisé en ce que si l'espace comporte au moins deux dimensions,

- l'extraction des grandeurs statistiques comporte un ajustement polynomial d'une courbe avec les points du nuage d'une région, et la représentation de cet ajustement polynomial par une suite de valeurs ajustées, et

10 - la définition du critère de ressemblance comporte une détermination d'un troisième seuil de comparaison pour comparer de telles suites de valeurs ajustées.

15 7. Procédé selon la revendication 6, dans lequel la fonction de similarité entre deux régions dépend de positions de centroïdes de nuages de deux régions, et d'extrémités de deux segments de courbe représentant ces deux régions.

8. Procédé selon la revendication 7, dans lequel la fonction de similarité est la distance entre les centroïdes des nuages de points.

20 9. Procédé selon l'une des revendications 6 à 8, dans lequel l'ajustement polynomial est une régression linéaire.

25 10. Procédé selon l'une des revendications 6 à 9, dans lequel les nuages de points décrivant une région sont distribués dans un espace dont trois dimensions de base sont trois combinaisons linéaires ou non linéaires distinctes des trois couleurs primaires de la synthèse additive et une autre dimension dans cet espace étant l'effectif des pixels selon cette distribution.

30 11. Procédé selon l'une des revendications 5 à 9, dans lequel les points d'un nuage décrivant une région sont distribués dans un espace dont trois dimensions de base sont la teinte, la saturation et l'intensité de la couleur, et une autre dimension dans cet espace étant l'effectif des

pixels selon cette distribution.

12. Procédé de segmentation selon la revendication 11 dans lequel le calcul de similarité entre régions est réalisé :

- en utilisant l'espace teinte/saturation/intensité si une saturation moyenne est supérieure à un quatrième seuil prédéterminé ;
- dans l'espace des intensités, si la saturation est inférieure ou égale à ce quatrième seuil.

13. Procédé selon l'une quelconque des revendications 1 à 9, dans lequel le signal électromagnétique apparaissant dans au moins deux images est transformé pour en extraire pour chaque région au moins deux composantes de mouvement entre les deux images pour le représenter, telles que valeur scalaire et orientation, et où les autres étapes du procédé s'appliquent à cette représentation.

14. Procédé de segmentation spatiale d'une image en objets visuels, caractérisé en ce que :

- a) au moins deux segmentations sont réalisées selon l'une des revendications 1 à 12 avec au moins deux tailles de régions initiales correspondant à des résolutions différentes ;
- b) on superpose les objets obtenus,
- c) on remplace le contour de l'objet obtenu par la résolution la moins fine par le contour obtenu par la résolution la plus fine, et/ou
- d) l'image est représentée comme une arborescence d'objets, chaque objet de niveau supérieur pouvant inclure un ou plusieurs objets de niveau inférieur.

15. Procédé de segmentation selon l'une quelconque des revendications précédentes dans lequel les seuils prédéterminés sont choisis pour maintenir la distribution des tailles des régions dans un intervalle défini pour éviter une sur et une sous-segmentation.

16. Procédé selon l'une des revendications 1 à 15, dans lequel des paramètres de seuil, de degré de polynôme et de fonction de similarité

sont choisis de façon adaptative par une méthode d'apprentissage prédéterminée, en fonction de seuils de sur-segmentation et de sous-segmentation à éviter, et d'un calcul d'évaluation prédéterminé de celles-ci.

5           17. Utilisation du procédé selon l'une des revendications 1 à 16, pour caractériser l'image par un ensemble d'objets décrits par plusieurs caractéristiques, l'une au moins de ces caractéristiques étant représentées par un nombre ou par une valeur dans un ensemble énuméré.

10           18. Utilisation du procédé selon l'une des revendications 1 à 17, pour classer les objets visuels de plusieurs images segmentées à l'aide de partition de l'espace de représentation des objets par une méthode de classification avec apprentissage choisie parmi une méthode de type neuronale ; une méthode de type génétique et une méthode d'analyse des données.

15           19. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 et d'une mesure de distance entre les courbes d'ajustement dans un espace de représentation de ces courbes, pour détecter la présence d'un même objet dans plusieurs images, consécutives ou non.

20           20. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 avec des méthodes statistiques de caractérisation des concepts contenus dans un texte, pour caractériser le contenu d'une suite de scènes portant sur un même sujet.

25           21. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 avec la reconnaissance d'un locuteur, pour caractériser le contenu de plans et de scènes d'une vidéo.

          22. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 avec un alignement automatique entre un script et une vidéo, pour caractériser le contenu des plans et des scènes d'un document vidéo.

30           23. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 avec un alignement automatique entre un script et une vidéo, pour

apprendre de façon automatique les caractéristiques des objets présents dans la vidéo et dans le script.

24. Utilisation conjointe du procédé selon l'une des revendications 1 à 17 avec la transcription de la parole de la vidéo sous forme de texte par  
5 une méthode de dictée vocale, pour caractériser le contenu de plans et de scènes d'une vidéo.

1/4

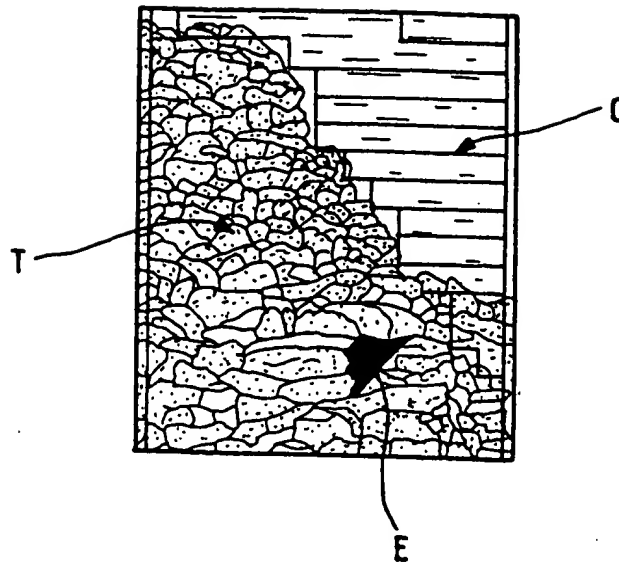


FIG. 1

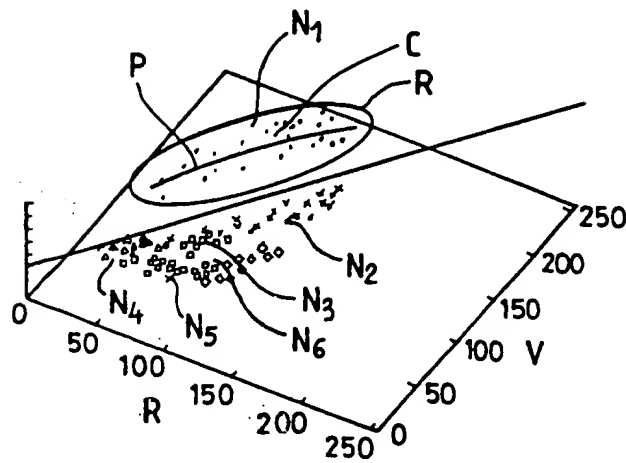


FIG. 2a

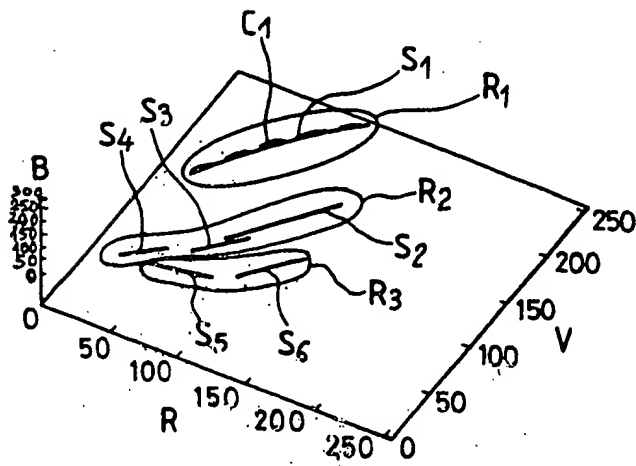
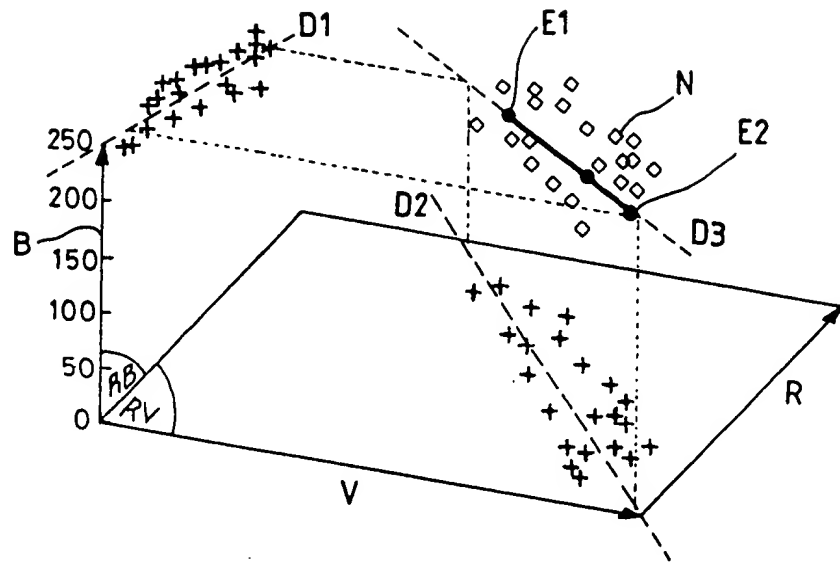
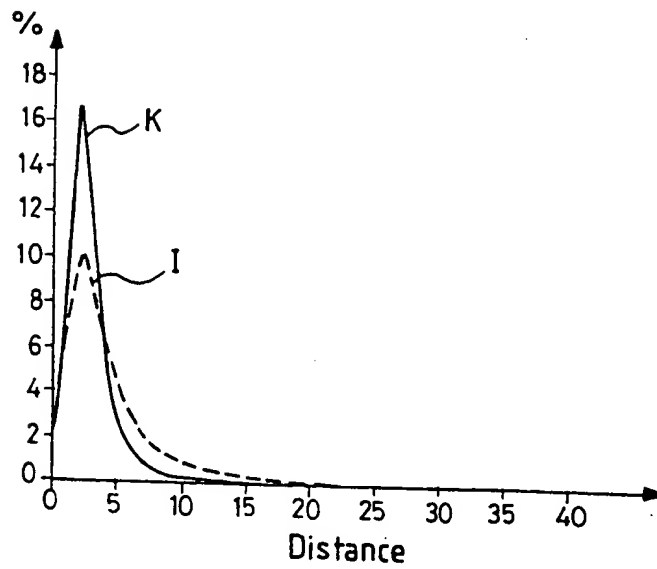


FIG. 2b

2/4



FIG\_3



FIG\_4

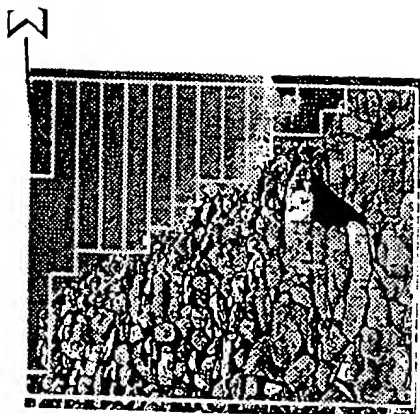


FIG 7a

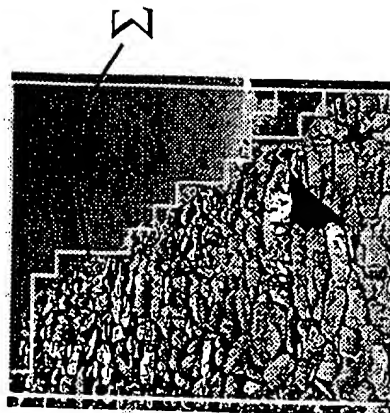


FIG 7b

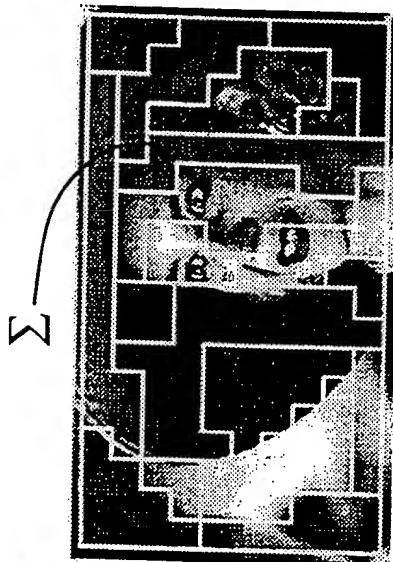


FIG 6a

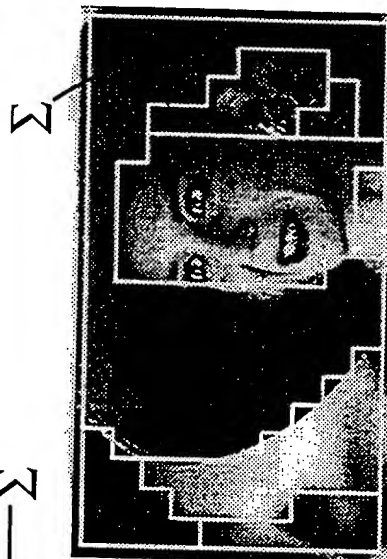


FIG 6b



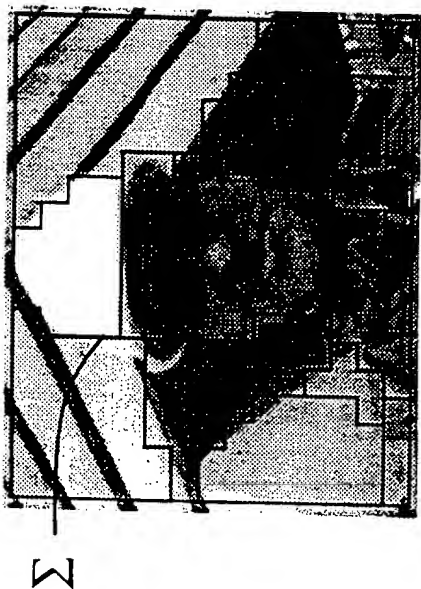
FIG 5a



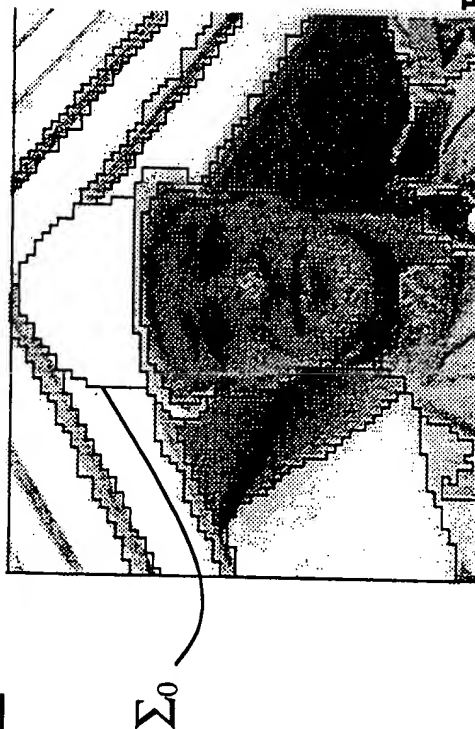
FIG 5b



**FIG 8b**



**FIG 8a**



**FIG 9**

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/FR 99/00176

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G06T5/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>SCHROETER P ET AL: "Hierarchical image segmentation by multi-dimensional clustering and orientation-adaptive boundary refinement" PATTERN RECOGNITION, vol. 28, no. 5, 1 May 1995, page 695-709 XP004011495 see abstract see page 697, left-hand column, line 25 - page 699, left-hand column, line 19 --- -/--</p>	1-24

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- document member of the same patent family

Date of the actual completion of the international search

31 May 1999

Date of mailing of the international search report

09/06/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Gonzalez Ordóñez, O

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 99/00176

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LI Y ET AL: "SEMANTIC IMAGE RETRIEVAL THROUGH HUMAN SUBJECT SEGMENTATION AND CHARACTERIZATION" STORAGE AND RETRIEVAL FOR IMAGE AND VIDEO DATABASES 5, SAN JOSE, FEB. 13 - 14, 1997, no. VOL. 3022, 13 February 1997, pages 340-351, XP000742394 SETHI I K;JAIN R C (EDS ) see abstract see paragraph 1 - paragraph 2 ---	1-24
A	EP 0 596 412 A (MATSUSHITA ELECTRONICS CORP) 11 May 1994 see abstract see page 4, line 54 - line 58; figures 2,3 -----	1-25

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Patent Application No

PCT/FR 99/00176

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0596412 A	11-05-1994	JP 6150000 A	31-05-1994
		DE 69324207 D	06-05-1999
		US 5519789 A	21-05-1996
<hr/>			

# RAPPORT DE RECHERCHE INTERNATIONALE

Dem. Internationale No  
PCT/FR 99/00176

<b>A. CLASSEMENT DE L'OBJET DE LA DEMANDE</b> CIB 6 G06T5/00				
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB				
<b>B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE</b> Documentation minimale consultée (système de classification suivi des symboles de classement) CIB 6 G06T				
Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche				
Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)				
<b>C. DOCUMENTS CONSIDERES COMME PERTINENTS</b>				
Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées		
Y	SCHROETER P ET AL: "Hierarchical image segmentation by multi-dimensional clustering and orientation-adaptive boundary refinement" PATTERN RECOGNITION, vol. 28, no. 5, 1 mai 1995, page 695-709 XP004011495 voir abrégé voir page 697, colonne de gauche, ligne 25 - page 699, colonne de gauche, ligne 19 --- -/--	1-24		
<input checked="" type="checkbox"/> Voir la suite du cadre C pour la fin de la liste des documents				
<input checked="" type="checkbox"/> Les documents de familles de brevets sont indiqués en annexe				
<b>* Catégories spéciales de documents cités:</b>				
<table border="0"> <tr> <td>           "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent            "E" document antérieur, mais publié à la date de dépôt international ou après cette date            "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)            "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens            "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée         </td> <td>           "T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention            "X" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément            "Y" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier            "Z" document qui fait partie de la même famille de brevets         </td> </tr> </table>			"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent "E" document antérieur, mais publié à la date de dépôt international ou après cette date "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée) "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée	"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention "X" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément "Y" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier "Z" document qui fait partie de la même famille de brevets
"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent "E" document antérieur, mais publié à la date de dépôt international ou après cette date "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée) "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée	"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention "X" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément "Y" document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier "Z" document qui fait partie de la même famille de brevets			
Date à laquelle la recherche internationale a été effectivement achevée 31 mai 1999		Date d'expédition du présent rapport de recherche internationale 09/06/1999		
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Fonctionnaire autorisé Gonzalez Ordenez, O		

Formulaire PCT/ISA/210 (deuxième feuille) (juillet 1992)

# RAPPORT DE RECHERCHE INTERNATIONALE

Don . Internationale No  
PCT/FR 99/00176

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	LI Y ET AL: "SEMANTIC IMAGE RETRIEVAL THROUGH HUMAN SUBJECT SEGMENTATION AND CHARACTERIZATION" STORAGE AND RETRIEVAL FOR IMAGE AND VIDEO DATABASES 5, SAN JOSE, FEB. 13 - 14, 1997, no. VOL. 3022, 13 février 1997, pages 340-351, XP000742394 SETHI I K; JAIN R C (EDS ) voir abrégé voir alinéa 1 - alinéa 2 ---	1-24
A	EP 0 596 412 A (MATSUSHITA ELECTRONICS CORP) 11 mai 1994 voir abrégé voir page 4, ligne 54 - ligne 58; figures 2,3 -----	1-25

### Renseignements relatifs aux membres de familles de brevets

PCT/FR 99/00176

Formulaire PCT/ISA/210 (annexe familles de brevets) (juillet 1992)